1. INTRODUCTION

1.1 Breast Cancer

Breast cancer has become a major health problem worldwide and in India. Breast cancer is mass of cells (tumour) that arise due to uncontrolled proliferation and growth of cells of breast tissue, either ductal or lobular. It occurs mainly in women with incidence rate 100 times more than that of male. Breast cancer may be in situ or invasive (malignant). Malignant are cancerous and can invade surrounding tissues and other organs of the body via lymph or blood. Breast cancer is a complex and miscellaneous disease. Many factors are involved in the pathogenesis of breast cancer (Ponder, 2001).

Breast cancer is a neoplastic condition that affects the breast tissue. Earlier, it was believed that breast carcinoma is a result of a process of histological changes in the mammary epithelium and is a multi-step process which starts with hyperplasia, progressing through atypical hyperplasia to in situ carcinoma and finally to invasive malignant carcinoma (Vogelstein and Kinzler 1998). With the advancement of molecular biology and genetics new reports and facts came into picture and nowadays, it is suggested that breast cancer is a complicated, polygenic diseases that involves various factors and multi-step processes (Beckmann et al., 1997; Ponder, 2001; Antoniou and Easton, 2006).

1.2 Epidemiology

1.2.1 Worldwide

Breast cancer is the common cause of cancer death in women worldwide. According to Globocan 2012 estimates, 522,000 breast cancer death cases were found in 2012 worldwide and are estimated to rise up to 13 million in
next 25 years. Approximately 1.7 million new cases of breast cancer were
diagnosed which are likely to increase to 42 million in next 25 years. The
incidence of breast cancer has increased much faster which is around 20%
since 2008 (Globocan, 2012). The number of young breast cancer cases
(between ages 15–49) is twice more in developing countries than developed
countries (Forouzanfar et al., 2011).
Breast cancer is the second most common cancer worldwide and it is the most
common cancer among women comprising 25% of all cancers. The estimated
new cancer cases diagnosed in 2012 are 1.67 million. Both in high and low
developed regions breast cancer is the most common cancer with more cases
in less developed (883,000 cases) than in more developed (794,000) regions.
Incidence rates vary in different regions of world, with rates ranging from
27/100,000 in Middle Africa and Eastern Asia to 92 in Northern America
(Globocan, 2012).
In case of cancer deaths, breast cancer ranks fifth with 522,000 deaths per
year. It ranks first as cause of cancer death in women in less developed
regions with 324,000 deaths (14.3%) per year. It is now the second most
common cause of cancer death in more developed regions with 198,000
deaths (15.4%) per year. The mortality rates ranges from 6/100,000 in
Eastern Asia to 20/100,000 in Western Africa and is lesser than that of
incidence because of the more survival of breast cancer in developed regions
(Globocan, 2012).
If we compare breast cancer in India with one of western country (USA) and
our neighbour (China), we can see that India has the maximum number of
female deaths due to breast cancer i.e.70,000 (Fig. 1.1).
Introduction and Review of literature

Molecular genetic study of breast cancer to identify susceptible genes in the population of West India

Figure 1.1 Estimated incidence, mortality and prevalence of breast cancer worldwide in a) 2008 and b) 2012 (Globocan, 2012)

In year 2012, the number of incidence of breast cancer in United States were 232,714 and female deaths were 43,909. So, in USA, one women is dying for
Introduction and Review of literature

Molecular genetic study of breast cancer to identify susceptible genes in the population of West India

every 5 or 6 women newly diagnosed. In China, the number of incidence of breast cancer were 187,213 and female deaths were 47,984. So, in China, one women is dying for every 4 women newly diagnosed. While for India, the number of incidence of breast cancer were 144,937 women and female deaths were 70,218. So, in India, one women is dying for every 2 women newly diagnosed. Though the incidences of breast cancer are higher in USA, yet the female deaths are lower. USA has decrease in mortality.

1.2.2 India

For approximately 4 decades, cervical cancer was the most common cancer in women in India and more deaths were attributed to cervical cancer than any other cancer. But over last one decade or so, breast cancer has been rising gradually, and now, breast cancer is the most common type of cancer in women in India. Both, the number of new cases, as well as female deaths, due to breast cancer are more than cervical cancer. The incidence of cervical cancer is decreasing while of breast cancer is increasing (Fig. 1.2 and 1.3).

![Incidence](http://example.com/incidence.png)
Introduction and Review of literature

Molecular genetic study of breast cancer to identify susceptible genes in the population of West India

Figure 1.2 Incidence of breast cancer in India a) in 2008 b) in 2012 (Globocan, 2012)

Figure 1.3 Number of new cases and deaths of breast cancer in India in 2012 (Globocan, 2012)
Increase in breast cancer cases is because of rapid rise in the numbers of breast cancer cases. The breast cancer burden in India has reached about 2/3rds of that of the US and is gradually rising.

1.2.2.1 West India

The increasing incidences of breast cancer cases show that it is a major threat to Indian society. Reports of ICMR show that the incidences of breast cancer have doubled in metropolitan cities from 1982 to 2005 (ICMR, 2010). It is found to be the most common type of cancer in women from Himachal Pradesh, Nagaland, Rajasthan, and Goa, and second most common type of cancer in women from Maharashtra, Gujarat and Punjab. Tripura is at third place in case of breast cancer (ICMR, 2001, Sharma et.al, 2009 and Gaur et al., 2006). This shows that the Western states (Rajasthan, Goa, Gujarat and Maharashtra) have highest number of breast cancer cases. The rise in the incidence of breast cancer is mainly in the premenopausal females of India (Agarwal, 2007).

1.3 Anatomy of breast

Development of breast development starts around about 7th-8th weeks after conception. Nipples, areola and ducts are formed at later stages of gestation. Breast tissues mature completely only after pregnancy. Female breasts are consist of specialized glands (lobules) that can produce milk. The mammary gland consists of around 15-20 lobes, each of which has a branching duct system ending in terminal ducts. Lobules are clusters of epithelial cells which originates from the nipple and ends in dozens of small bulbs like structures that can produce milk and are called alveoli. The milk travels through a network of tiny tubes (ducts) to alveoli. They are connected
together by dense connective tissue septa. The contour of the breast is filled out by fat tissue. Areola are the dark brown area that surrounds the nipples (Fig. 1.4). The breast also contains blood & lymph vessels and lymph nodes (Boron and Boulpaep, 2003).

1.3.1 Lymphatic System of Breast

The lymphatic system is one of the major paths through which breast cancer spreads. Lymphatic system consists of lymph vessels which carry a clear fluid called lymph which drains into lymph nodes. Lymph nodes are small bean-shaped structures which contain cells involved in immune system.
Introduction and Review of literature

Molecular genetic study of breast cancer to identify susceptible genes in the population of West India

b)

Figure 1.4 Women breast anatomy a) cross section, b) enlarged view of lobule and alveolus (Source: Boron and Boulpaep, 2003 and Halperin et al., 2008)

Lymph is drained by lymph vessels into

I. Supraclavicular and infraclavicular lymph nodes around the collar bone.

II. Axillary nodes that are present under the arm.

III. Internal mammary nodes that are present inside the chest near the breast bone.

1.4 Types of breast cancer

It is divided into a number of histological subtypes i.e., ductal which arises in milk ducts, lobular which arises in lobules of breast tissue, tubular, medullary and papillary cancer (Fig. 1.5). These are further classified as in situ and invasive (Lishman and Likhani, 1999; and Mai et al., 2000). The most common subtype is invasive ductal carcinoma (80–90%) followed by invasive lobular carcinoma.
(approximately 10%). Medullary (3–5%) and papillary subtypes are less common (Ellis et al, 1992).

**Figure 1.5** Types of breast cancer

The breast tumours depict different growth characteristics. Thus, these can be distinguished on the basis of histological grading (tumour markers) used
clinically in classification of breast cancer. These markers include estrogen receptor (ER), progesterone receptor (PR) and Human Epidermal Growth Factor Receptor type 2 (HER2) status. The prognosis of disease is defined on the basis of these clinical markers. Also, the molecular subtypes based on gene expression profiling have been established (Perou et al, 2000). Breast cancer in premenopausal and postmenopausal women also show totally distinct characteristics and require different treatment strategies (Schottenfield and Joseph, 1996).

**Molecular subtypes of breast cancer:**

- **Luminal A:** This type of breast cancers are estrogen-receptor (ER) and/or progesterone-receptor (PR) positive and HER2 negative. They are low- class, develop slowly and have easy treatment. They have low levels of Ki-67.

- **Luminal B:** This type of breast cancer is ER and/or PR positive and HER2 positive or negative. They have high levels of Ki-67. They develop faster than luminal A and their treatment is moderately worse.

- **Triple-negative/basal-like:** This type of breast cancer is ER, PR and HER2 negative. It is more common in young females who have *BRCA1* gene mutations.

- **HER2-enriched:** This type of breast cancer is ER/PR negative and HER2 positive. They develop faster than luminal cancers and have bad treatment.

- **Normal-like:** This type of breast cancer is almost similar to luminal A type of breast cancer but has moderately worse treatment process than it (Lishman and Likhani, 1999; and Mai et al., 2000).
Breast cancer types on the basis of tumor location:

On the basis of location breast cancer is divided into two types:

1. **In situ**: It remains to its place of origin.
   
   i. Ductal carcinoma in situ (DCIS): It is the most frequent type of breast cancer. It originates inside the milk ducts, remains in its original place only. It is non-invasive and less dangerous but women with this type of cancer are at higher risk recurrence. It comprises almost 80% of all breast cancer.

   ii. Lobular carcinoma in situ (LCIS): It is a type of breast cancer in which cancer cells start growing in the lobules (milk-producing glands). The cancer cell growth remains confined to the lobules and does not spread to surrounding tissues. LCIS is not a real breast cancer. LCIS is more common in women with ages between 40 and 50.

2. **Invasive**: It spreads to the surrounding area from its origin.

   i. Invasive lobular carcinoma (ILC): It is also called infiltrating lobular carcinoma. This type of cancer starts from lobule walls and spread to surrounding tissues. It ranks second in breast cancer after invasive ductal carcinoma. Nearly 10% of all invasive breast cancers are invasive lobular carcinomas. It is present in women of all ages but more common in older.

   ii. Invasive ductal carcinoma: This of cancer originates from the wall of milk ducts and invaded or spread to the surrounding breast tissues. These can spread to the lymph nodes and may also to other parts of the body.
Invasive ductal carcinoma is of five types namely:

- **Tubular carcinoma:** It is a subtype of invasive ductal carcinoma. These are small and made up of tube-shaped structures called "tubules." These are low-grade, less aggressive and constitute about 1-4% of all breast cancers. Now that screening mammography is widely used, however, tubular carcinomas are being diagnosed more frequently — often before you or your doctor would be able to feel a lump. It is present in women of all age group but most common in females in early 50s.

- **Medullary carcinoma:** This type of cancer is soft and similar to brain medulla. It starts with milk duct and spread nearby. It is rare and constitute 3-5% of all breast cancer cases. It occurs at any age but affects women of age between 40 and 50. It resembles aggressive cancer but behaves as normal. It doesn’t spread outside the breast to the lymph nodes.

- **Mucinous carcinoma:** is also called colloid carcinoma: It is a rare carcinoma in which cancer cells float in pools of mucus. It constitutes only about 2-3% of invasive breast cancers. It is present in all age group but mostly found in women in 60s or early 70s. It is less aggressive type and have less chance of spreading to the lymph nodes than other types of breast cancers.
**Introduction and Review of literature**

- **Papillary carcinomas:** This type of cancer is invasive, small and have finger-like projections. This constitutes less than 1-2% of invasive breast cancers and is grade 2 type.

- **Cribriform Carcinoma:** In this type of carcinoma, the cancer cells invade the stroma/fluid between the ducts and lobules. In this type of cancer, cells have holes on surface. It is usually low grade and constitute about 5-6% of invasive breast cancers (Ellis et al, 1992).

**1.5 Breast cancer risk factors**

Breast cancer is a complex disease which involves multiple pathways and factors. These risk factors include genetic and non-genetic factors (Fig. 1.6).

1.5.1 Non-genetic

Non-genetic factors along with genetic factors play an important role in the development of breast cancer.

1.5.1.1 Age

It is reported that women with older age (above 50) years are more susceptible to develop this disease (Feuer et al., 1993) i.e., the risk of breast cancer increases with increase in age. But in last one decade, the incidence of breast cancer occurrence in young women is found to be more in developing countries like India, Pakistan etc. (Kakarala et al., 2010).

1.5.1.2 Menstrual and reproductive history

Women who have early menarche (at age <12 years) and late menopause (age >55 years) are more prone to develop this disease. Also, females who have late pregnancy (after 30 years) and no pregnancy or/and who do not
breastfeed or breastfeed for smaller period are at higher risk (Singletary, 2003).

1.5.1.3 Hormone therapy

Women who had undergone prolonged hormone replacement therapy are 20% more likely to develop this cancer. It is been observed that more they are exposed to estrogen or progesterone hormones, more is the risk of developing breast cancer (Singletary, 2003).

1.5.1.4 Mammographic density

Females with high mammographic density are at 5 fold higher risk than women with less density (Boyd et al., 1995).

1.5.1.5 History of other benign breast disease

Women who had history of benign breast diseases like fibrocystic disease (fibro-adenosis), atypical hyperplasia, fibro-adenoma etc. have higher risk of developing breast cancer (Sabiston and Lyerly, 1997). Those with atypical hyperplasia history have 3–5 times more risk than women without any breast disease (Carter et al., 1988).

1.5.1.6 Exposure to radiation

X-rays and UV rays are known to damage genomic DNA in many ways. Exposure to radiations like X-rays at younger age (during puberty) increases the risk of developing breast cancer (John and Kelsey, 1993).

1.5.1.7 Ethnicity

Ethnic difference is also a factor that affects breast cancer prevalence. For example, breast cancer is more common among whites in USA. Ethnic differences in estrogen receptor (ER) and progesterone receptor (PR) are found to affect the probability of breast cancer (Uptodate, 2015). In a
Introduction and Review of literature

Multiethnic Cohort Study, various statuses of ER/PR have been reported and ER/PR status varied significantly across racial/ethnic groups even within the same tumor stage. African-American women have high prevalence of hormone receptor-negative tumors compared to whites, which may contribute to their high breast cancer mortality (Setiawan et al., 2009).

---

**Figure 1.6 Breast cancer risk factors**

- **Non-genetic**
  - Non-modifiable:
    1. Older age
    2. Early menarche
    3. Late menopause
    4. Late pregnancy
    5. High mammographic density
    6. History of benign breast disease

- **Modifiable**
  1. Life style: smoking, less physical activity, alcohol, late/no marriages, obesity
  2. Hormone replacement therapy
  3. Exposure to radiation

- **Genetic**
  - **High-penetrance genes**:
    1. Family history of breast cancer
    2. Mutation of BRAC1/2, ATM, p53 gene
  - **Low-penetrance genes**:
    Genetic variants in various genes e.g., genes involved in estrogen and carcinogen metabolism, apoptosis, DNA repair etc.
Introduction and Review of literature

1.5.1.8 Lifestyle

Lifestyles of individuals greatly influence the risk of breast cancer. More westernized lifestyle (less physical activities, higher fat diets, obesity, intake of alcohol and smoking) enhance breast cancer risk. Physical activity is considered protective because it reduces ovulatory cycle and increases levels of other form of estrogen i.e. catechol-o-methylated estrogens (Henderson et al., 1985). High fat diets have high level of cholesterol which is a precursor in estrogen synthesis (Aguas et al., 2005). Obesity has been found to be associated with an increase in estrogen levels (Hunter and Willet, 1993). Cigarette smoking produces potent carcinogens (Mitruen and Hirvonen, 2003). Alcoholic women are more prone to develop breast cancer because they have higher levels of estrogens than non-alcoholic (Kuper et al., 2000 and Reichman et al., 1993).

1.5.2 Genetic Factors

Even though some women follow westernized style of living yet they may not develop the breast cancer whereas some who do not follow westernized style develop breast cancer. This is where the genetic factors play their role. Genetic factors can modify the risk of breast cancer development (Lichtenstein et al., 2000).

1.5.2.1 Family history

Breast cancer cases with history of cancer in family accounts for 10–15% of all breast cancer cases (Loman et al., 2001). The risk is 80% if the first degree relative was diagnosed breast cancer at age >50 years while it is 330% if diagnosed at premenopausal age (Singletary, 2003). History of ovarian cancer also increases the risk of breast cancer (Mahoney et al., 2008).
1.5.2.2 High penetrance genes

Mutations or rare allelic variants of high penetrance genes i.e., BRAC1/2, tumor protein 53 gene (p53) and ataxia telangiectasias mutated gene (ATM) confer high breast cancer risk. BRAC1/2 mutation associated breast cancer cases account for about 2–3% of all cases (Narod and Foulkes, 2004).

1.5.2.3 Low penetrance genes

It is known that genetic factors play important role in breast cancer pathogenesis. Inherited high risk/high penetrance genes constitute only about 5-10% of total breast cancer cases. A large number of studies related to genetics of breast cancer show that it is a polygenic diseases which involves many more genes that are unknown. These genes confer low risk to breast cancer. But together with other factors like lifestyle and environmental, this low risk may combine and enhance the risk of developing this diseases. Low penetrance genes involve the gene coding enzymes responsible for hormone metabolism (mainly estrogen), carcinogens metabolism, detoxification of reactive oxygen species and DNA repair (Mitruen and Hirvonen, 2003). These genes confer low risk of breast cancer and may modify the risk through polygenic mechanism. To study these genes many approaches are used, of which the association studies of variants with breast cancer are mostly used (Le and Wilkens, 2008).

1.6 Breast Cancer Diagnosis and Treatment in India

Diagnostic methods available are clinical observation, mammography and tissue biopsy. Treatment plan of breast cancer includes:

i. Surgery: It involves removal of lump or surrounding tissue (lumpectomy), complete breast except muscles (mastectomy) and lymph nodes (Abeloff et al., 2008 and Institute NC, 2011).
ii. Radiation therapy: It includes external beam of radiation given to the patient to remove the remaining cells left after surgery (Abeloff et al., 2008 and Institute NC, 2011).

iii. Chemotherapy: It is given to reduce or destroy cancer cells. It may be given before surgery (to reduce the size of tumor) or after surgery (to reduce the recurrence risk) (Abeloff et al., 2008 and Institute NC, 2011).

iv. Hormone therapy: This treatment is given to the breast cancer patients who are hormone receptor positive. It involves either the blockage of synthesis of estrogen (removal of ovaries, aromatase inhibitors etc.) or the reduction of estrogen amounts in the body (anti-estrogens e.g., Tamoxifen) (EBCTCG, 2005).

1.7 Genetics of breast cancer

Each individual have their own genetic makeup and show different responses to life style or environmental factors. In addition to high penetrance genes, a combination of high-risk variants in low- and medium-penetrance genes also confer genetic predisposition to cancer. The prognosis may differ depending on their genetic combination. Thus, it is important to consider the genetic variants in relation to the other risk factors to detect the combined effect on breast cancer risk (Le and Wilkens, 2008).

1.8 DNA repair system

Human cells are continuously exposed to endogenous and exogenous agents that damage genetic material i.e. DNA (Hakem, 2008). Unrepaired DNA affects genomic stability (Friedberg, 2002; Sengupta and Harris, 2005). The variants generated because of mutations arising due to these DNA damages can confer alterations in cell cycle, apoptosis and can cause genomic instability. This genomic instability can further lead to cellular transformation to cancer cell (Govan et al., 1990) (Fig. 1.7). Normally three mechanisms are involved in blocking mutagenesis due to damaged DNA by rectifying the DNA damage or
Introduction and Review of literature

Molecular genetic study of breast cancer to identify susceptible genes in the population of West India

activating some response: 1) DNA repair system 2) Cell-cycle checkpoint control and 3) Apoptosis (Branzei and Foiani, 2008). DNA repair systems repair the damages generated in DNA while apoptotic system destroys cells that have been subjected to irreparable DNA damage. DNA repair systems are the most effective cellular systems countering DNA damages because they restore the DNA structure.

Figure 1.7 How DNA repair impairment leads to development of cancer cell

Cellular DNA repair mechanisms in eukaryotes consists of at least five mechanisms: a) Base excision repair (BER) b) Nucleotide excision repair (NER) c) Mismatch repair (MMR) d) Homologous recombinational repair (HRR) and e) Non-Homologous End-joining (NHEJ) pathway (Painter 1978) (Fig. 1.8).

BER is a crucial repair system against endogenous DNA damages resulting from
cellular metabolism including lesions arising from reactive oxygen species (ROS), methylation, deamination and hydroxylation, and other lesions generated from other exogenous agents including chemicals, radiations etc. (Zharkov, 2008). BER focuses on small chemical alterations of the DNA bases, which creates abrasions (Hegde et al., 2008). These abrasions causes difficulties in normal transcription and replication. These abrasions are highly deleterious to genomic stability as they may cause increase in spontaneous mutations (Weissman et al., 2007).

![DNA repair pathways and their role](image)

**Figure 1.8** DNA repair pathways and their role. Modified from (Boland et al., 2005)
1.8.1 Base Excision Repair pathway

BER is initiated by DNA glycosylases by recognizing the damaged base (Fig. 1.9).

![Base Excision Repair Pathway](image)

**Figure 1.9** BER pathway (Singh and Mistry, 2017)

These release the damaged base resulting in an apurinic/apyrimidinic site. This site is then cleaved by APE1, leaving a 5’-deoxyribose phosphate residue which is removed by DNA polymerase β. DNA polymerase β also
inserts new correct nucleotide. DNA ligase III then finally seals the gap between the strands of DNA. The X-ray repair cross-complementing group 1 (XRCC1) coordinates the activity of DNA polymerase-β and DNA ligase to fill the gap (Li et al., 2011).

1.8.1.1 *NEIL1* gene

DNA glycosylases are the first enzyme to initiate the BER pathway. NEIL1 is one of the major DNA glycosylase which can cleave lesions present in single-stranded, double-stranded, and bubble DNA structures (Dou et al., 2003). The human *NEIL1* gene is located on chromosome 15q23, and mutations in *NEIL1* are found to cause increased risk of primary gastric cancer in a Japanese population (Shinmura et al., 2004). *NEIL1* gene polymorphisms have also been studied for association with other cancer risk (Table 1.1).

**Table 1.1** Case-control studies of polymorphisms in *NEIL1* gene

<table>
<thead>
<tr>
<th>Reference</th>
<th>SNP</th>
<th>Population</th>
<th>Cancer type</th>
<th>No. of cases</th>
<th>No. of controls</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Goto et al., 2010</td>
<td>c.3769C&gt;T, c.3170T&gt;G</td>
<td>Japan</td>
<td>Gastric</td>
<td>85</td>
<td>-</td>
<td>Positive correlation with gastric cancer</td>
</tr>
<tr>
<td></td>
<td>c.2681TA[8]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Han, 2009</td>
<td>-</td>
<td>Caucasians (USA)</td>
<td>Breast</td>
<td>239</td>
<td>477</td>
<td>No significant association</td>
</tr>
<tr>
<td>Zhai et al., 2008</td>
<td>rs7182283 rs4462560</td>
<td>American</td>
<td>Oral carcinoma</td>
<td>872</td>
<td>1044</td>
<td>No effect</td>
</tr>
</tbody>
</table>

1.8.1.2 *XRCC1* gene

XRCC1 protein is an important DNA repair proteins of BER pathway. It is a multidomain protein which plays an important role by interacting with DNA polymerase β, APE1, poly(ADP-ribose) polymerase (PARP-1), and...
DNA ligase III and has no known catalytic domain (Bu et al., 2006). It is located on chromosome 19q13.2~13.3, contains 17 exons, spans approximately 31.9 kb and encodes for a 633 amino acid protein (Zhenga et al., 2009 and Liu et al., 2009).

X-ray repair cross-complementing 1 protein (XRCC1; locus=chromosome 19q13.2) is a crucial protein of base excision repair pathway (BER) which repairs the most frequent DNA aberrations i.e., single strand break (Zhenga et al., 2009; Bu et al., 2006 and Liua et al., 2009). It also repairs damages caused by oxidative stress and exogenous sources like chemicals, UV and ionizing radiations (Hirata et al., 2007). If these damages remain in the genome, they may lead to the transformation of normal cell to cancerous cell (Liu et al., 2011). Thus, this protein has important role in carcinogenesis. Various reports depicts that XRCC1 is a one of the low penetrance gene for cancer susceptibility. It is reported that XRCC1 deficient mouse are hypersensitive to ionizing radiations and its knockouts are lethal (Ladiges, 2006). It is also found to participate in non-homologous end joining (NHEJ) pathway and nucleotide excision repair (NER) that repairs damages induced by UVC (Audebert et al., 2004; Moser et al., 2007; Audebert et al., 2006).

Several studies have been reported which show that the polymorphism of XRCC1 gene can modify breast cancer risk (Liu et al., 2011, Chacko et al., 2005 and Li et al., 2009. Many SNPs have been studied for breast cancer risk, of which three SNPs (i.e. Arg194Trp (rs1799782), Arg280His (rs25489) and Arg399Gln (rs25487)) are widely studied. But the results are inconsistent (Table 1.2).
**Table 1.2** Case-control studies of polymorphisms in *XRCC1* gene

<table>
<thead>
<tr>
<th>Reference</th>
<th>SNP</th>
<th>Population</th>
<th>No. of cases</th>
<th>No. of controls</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patel et al., 2005</td>
<td>1) rs1799782</td>
<td>USA</td>
<td>485</td>
<td>485</td>
<td>Increased risk of breast cancer with rs25487 breast cancer</td>
</tr>
<tr>
<td></td>
<td>2) rs25487</td>
<td></td>
<td>452</td>
<td>452</td>
<td></td>
</tr>
<tr>
<td>Bu et al., 2006</td>
<td>1) rs25487</td>
<td>Black, Jewish, Caucasian, Hispanic, Ashkenazi</td>
<td>190</td>
<td>95</td>
<td>Increased risk of breast cancer with IVS10+141G&gt;A</td>
</tr>
<tr>
<td></td>
<td>2) IVS10+141G&gt;A</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Li et al., 2009</td>
<td>1) rs1799782</td>
<td>Mixed (Caucasian, Asian, USA)</td>
<td>10,465</td>
<td>10,888</td>
<td>rs25489 and rs25487 showed increased risk of breast cancer in Asian population only</td>
</tr>
<tr>
<td></td>
<td>2) rs25489</td>
<td></td>
<td>6156</td>
<td>5806</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3) rs25487</td>
<td></td>
<td>21467</td>
<td>22,766</td>
<td></td>
</tr>
<tr>
<td>Liu et al., 2011</td>
<td>277T&gt;C</td>
<td>Chinese</td>
<td>995</td>
<td>1004</td>
<td>Increased risk of breast cancer</td>
</tr>
<tr>
<td>Smolarz et al., 2014</td>
<td>rs1799782</td>
<td>Poland</td>
<td>70</td>
<td>70</td>
<td>No significant association with breast cancer</td>
</tr>
</tbody>
</table>

### 1.8.2 Non-Homologous End Joining pathway

NHEJ is considered to be the major repair pathway of DSBs in eukaryotic cells during most phases of the cell cycle (Pfeiffer et al., 2004). NHEJ involves many proteins namely, Ku70/80, DNA-dependent protein kinase catalytic subunit (DNA-PKcs), Artemis, XLF, X-Ray Repair Cross-Complementing Protein 4 (XRCC4), DNA ligase 4, ATM, p53 and MDM2 proteins (Bassing et al., 2002 and Shrivastav et al., 2008). Cells devoid of NHEJ pathway have increased genomic instability, thus, are more prone to tumorigenesis (Ferguson and Alt, 2001; and Gao et al., 2000) (Fig. 1.10).
Introduction and Review of literature

Molecular genetic study of breast cancer to identify susceptible genes in the population of West India

1.8.2.1 XRCC4

XRCC4 plays a crucial role in the NHEJ pathway. XRCC4 is responsible for accurate end-joining of DNA double strand breaks in fibroblasts of mammals (van Heemst et al., 2004) and their mouse model with inactivated gene showed late embryonic lethality in lymphocytes. This clearly shows that neurons and lymphocytes require XRCC4 for double strand break repair. It is located at position 5q14.2 on chromosome 5 and codes for 336 amino acids (Genecards; Entrez Gene id: 7518). Many studies were reported which investigated the polymorphism of XRCC4 and breast cancer risk. SNP (rs2075685) which located in the intron of XRCC4 gene was
found to be significantly associated with breast cancer in a northern Taiwan population (OR=0.583, P = 0.02) while SNPs rs1805377 and rs2075686 were not associated with breast cancer risk (Fu et al., 2003). Similar studies of different SNPs were reported in different populations which showed variable results (Table 1.3).

Table 1.3 Case-control studies of polymorphisms in XRCC4 gene

<table>
<thead>
<tr>
<th>Reference</th>
<th>SNP</th>
<th>Population</th>
<th>No. of cases</th>
<th>No. of controls</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fu et al., 2003</td>
<td>rs2075685</td>
<td>Taiwanese</td>
<td>254</td>
<td>379</td>
<td>S^a</td>
</tr>
<tr>
<td></td>
<td>rs1805377</td>
<td></td>
<td></td>
<td></td>
<td>NS^b</td>
</tr>
<tr>
<td></td>
<td>rs2075686</td>
<td></td>
<td></td>
<td></td>
<td>NS</td>
</tr>
<tr>
<td>Lee et al., 2005</td>
<td>rs1056503</td>
<td>Korean</td>
<td>872</td>
<td>671</td>
<td>NS</td>
</tr>
<tr>
<td>Garcia et al., 2006</td>
<td>rs1805377</td>
<td>Caucasian (Poland and USA)</td>
<td>10979</td>
<td>10423</td>
<td>NS</td>
</tr>
<tr>
<td>Allen-Brady et al., 2006</td>
<td>rs1478485</td>
<td>North European</td>
<td>457</td>
<td>576</td>
<td>S</td>
</tr>
<tr>
<td></td>
<td>rs13180316</td>
<td></td>
<td>458</td>
<td>576</td>
<td>S</td>
</tr>
<tr>
<td></td>
<td>rs963248</td>
<td></td>
<td>455</td>
<td>574</td>
<td>NS</td>
</tr>
<tr>
<td></td>
<td>rs1056503</td>
<td></td>
<td>459</td>
<td>576</td>
<td>NS</td>
</tr>
</tbody>
</table>

^aS-significant, ^bNS-nonsignificant

1.9 Estrogen receptor 1 (ESR1) gene

It is well established that prolonged estrogen exposure is a strong risk factor because of its activity to induce mammary epithelial cells to proliferate and grow. Its activity is mediated by two receptors belonging to nuclear receptor family, namely, ESR1 and ESR2. ESR1 has major role in estrogen signalling and is associated with breast cancer (Roodi et al., 1995 and Khan et al., 1998). ESR1 is located at chromosome 6p.25.1 and spans about 300 kb with eight exons (Li et al., 2010). It has one estrogen binding domain and one DNA binding domain. The DNA binding domain is hormone response element that binds to
**Introduction and Review of literature**

DNA and regulates the downstream expression of genes involved in cell proliferation and growth (Potter et al., 1995 and Rayter, 1991). Genetic variants of *ESR1* gene might therefore have important effects in breast carcinogenesis. Polymorphisms may increase estrogen-associated breast cancer risk.

**Table 1.4** Case-control studies of polymorphisms in *ESR1* gene

<table>
<thead>
<tr>
<th>Reference</th>
<th>SNP</th>
<th>Population</th>
<th>No. of cases</th>
<th>No. of controls</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parl et al., 1989</td>
<td>rs2234693, rs9340799</td>
<td>American, African-American</td>
<td>59</td>
<td>-</td>
<td>No effect</td>
</tr>
<tr>
<td>Yaich et al., 1992</td>
<td>rs2234693</td>
<td>American</td>
<td>257</td>
<td>140</td>
<td>No effect</td>
</tr>
<tr>
<td>Andersen et al., 1994</td>
<td>rs2234693, rs9340799</td>
<td>Norwegian Caucasian</td>
<td>360</td>
<td>672</td>
<td>Association with increased risk of breast cancer with G allele</td>
</tr>
<tr>
<td>Roodi et al., 1995</td>
<td>rs1801132, c.729C→T</td>
<td>American</td>
<td>188</td>
<td>-</td>
<td>rs1801132 associated with family history of breast cancer</td>
</tr>
<tr>
<td>Schubert et al., 1999</td>
<td>rs1801132, c.729C→T</td>
<td>Caucasian-American and African-American</td>
<td>105</td>
<td>151</td>
<td>No increased risk of breast cancer</td>
</tr>
<tr>
<td>Cai et al., 2003</td>
<td>rs2234693, rs9340799</td>
<td>Chinese</td>
<td>1069</td>
<td>1166</td>
<td>Increased risk with rs2234693</td>
</tr>
<tr>
<td>Shin et al., 2003</td>
<td>rs2234693, rs9340799</td>
<td>Korean</td>
<td>205</td>
<td>205</td>
<td>Decreased risk with rs9340799</td>
</tr>
<tr>
<td>Wang et al., 2007</td>
<td>rs746432, rs2234693, rs9340799, rs1801132</td>
<td>Caucasian USA</td>
<td>393</td>
<td>790</td>
<td>No effect of any of the SNP</td>
</tr>
<tr>
<td>Tsezou et al., 2008</td>
<td>-1174(TA)←27 Repeat</td>
<td>Greek</td>
<td>79</td>
<td>155</td>
<td>No effect</td>
</tr>
<tr>
<td>Dunning et al., 2009</td>
<td>rs3798577, rs2228480, rs1801132, rs9340799, rs2234693</td>
<td>Caucasians (United Kingdom)</td>
<td>2,276</td>
<td>2,188</td>
<td></td>
</tr>
</tbody>
</table>
Molecular genetic study of breast cancer to identify susceptible genes in the population of West India

The polymorphism of ER protein in ER positive breast cancer may modify the response to hormone treatment (Donegan, 1992). The *ESR1* was first linked to breast cancer by Zuppan et al. in 1991 (Zuppan et al., 1991). Later, Zheng et al. reported a strong association between SNP rs2046210 and breast cancer (Zheng et al., 2009). Mutations in coding region are found only in small percent of breast cancer patients (Southey et al., 1998). Variants in non-coding region are more common and widely studied. SNP rs2234693 (PvuII) was found to be altering the binding of transcription factor and increased expression of ESR1 (Parl et al., 1989). After that several studies have been carried out to detect the association of polymorphism of *ESR1* gene with breast cancer but results were inconsistent (Table 1.4).

1.10 Single nucleotide polymorphisms (SNPs)

The most commonly studied form of genetic variation are single nucleotide polymorphisms (SNPs). Polymorphism word is taken from Greek which means "having multiple forms". SNPs are commonly occurring (>1%) single base pair genetic mutations in at least one population. SNPs comprises of approximately 90% of all sequence variants in the human genome (Collins et al., 1998). Total SNPs found in human genome are about 11 million (Kruglyak and Nickerson, 2001).

A large number of non-synonymous single nucleotid polymorphisms (nsSNPs) or polymorphisms are identified in the human gene sequence. These mutant alleles present in the conserved regions of the candidate genes affect the capacity of the protein to work properly (Saadat et al., 2010). Genetic variants present in different regions affect the activity of protein differently. SNPs in exonic or coding regions may affect the sequence of protein and thus have impact on
Introduction and Review of literature

structure as well as function. SNPs present in intronic regions may affect the RNA splicing by disrupting or creating RNA splicing sites. Variants in 5’UTR and 3’UTR regions may disrupt the binding of miRNAs which carry out gene silencing or protein degradation. Variants in promoter regions affect transcription level by affecting the binding of transcription factor on the promoter site. The variants in conserved regions may affect the binding proteins with their respective ligands. The changes in the structure caused by the nsSNPs alter the interactions between protein and the respective ligands (Tebbs et al., 1999 and Masson et al., 1998).

Over the last one decade, many association studies has been conducted for searching low penetrance breast cancer susceptibility variants. These association studies for breast cancer involved testing functional SNPs in candidate genes, involved in important biological pathways such as DNA repair, carcinogen metabolism, cell-cycle control and hormone synthesis and metabolism (Pharoah et al., 2004).

Various association studies reported to found association of variants with breast cancer but are not replicable. These show different results and may give false positive results (Hirschhorn et al., 2002; Ioannidis et al., 2001; Ioannidis et al., 2006; Lohmueller et al., 2003; Wacholder et al., 2004). This lack of replication is mainly due to two factors: insufficient power of individual studies to detect small contributions to risk and low prior probability of a disease association for a given variant (Chanock et al., 2007). Recently, researchers from all over the world have joined together by collaborating with many groups and combining data from large number of breast cancer cases and controls from different ethnics groups, in order to overcome limitations of individual studies and have
Introduction and Review of literature

sufficient power to detect small effects on breast cancer risk (Breast Cancer Association Consortium, 2006).

So far, candidate gene approach on breast cancer have studied only a small fraction of the around 30,000 genes across the human genome. Of those studied genes only limited genetic variation of these genes have been evaluated. So, there is scope of identifying additional breast cancer susceptibility loci by candidate gene approach, in large population studies, better evaluation of genetic variation. This approach may also help in identifying rare variants in candidate genes which are associated with breast cancer risk (Garcia-Closas and Chanock, 2008).

1.11 In silico study

Studying effect of large number of SNPs in biological system is difficult task. Computational approach is easier and effective as compared to experimental approach to perform such studies. These analysis require either structure or sequence of protein to determine the phenotypic changes caused by substitution of new amino acids due to the presence of variants. The computational analysis is cost effective, reliable and require very less time in comparison to experimental analysis. Numerous in silico analysis are available that analyse the physicochemical properties of the wild type and mutated protein and predict the effects of SNP on structure and function of protein. Docking analysis detects the effect of variants on binding interactions of proteins with their respective ligands. Energy minimization studies evaluate the effects of SNPs at atomic level and on overall free energy level i.e. stability of the protein (Mah et al., 2011 and Brunham et al., 2005).
1.12 Purpose of study/Hypothesis

There is no way to prevent breast cancer, but it can be controlled by detecting it early and treat adequately. This is the only way to increase the survival rate of patients with breast cancer. Genetic polymorphisms are found to be responsible for predisposition of cancer. These can modify the risk associated with breast cancer and response of individuals to the treatment. Many evidences mentioned earlier show that the genetic variants of low penetrance genes have effect on breast cancer development. Several SNPs of targeted genes have been reported in dbSNP database but few are studied for their association with breast cancer. Association studies of SNPs of these low penetrance genes with breast cancer diseases are not found to be reported in the population of West India. Thus, these SNPs can be studied in this population to find out the association of breast cancer with the variant alleles in the population which could help in better prognosis of the disease.